

Article

Improving Wind Power Forecasts: Combination through Multivariate Dimension Reduction Techniques

Marta Poncela-Blanco ^{1,*} and Pilar Poncela ²¹ Joint Research Centre, European Commission, 21027 Ispra, Italy² Department of Análisis Económico, Economía Cuantitativa, Universidad Autónoma de Madrid, Avenida Tomás y Valiente 5, 28049 Madrid, Spain; pilar.poncela@uam.es

* Correspondence: Marta.PONCELA-BLANCO@ec.europa.eu

Abstract: Wind energy and wind power forecast errors have a direct impact on operational decision problems involved in the integration of this form of energy into the electricity system. As the relationship between wind and the generated power is highly nonlinear and time-varying, and given the increasing number of available forecasting techniques, it is possible to use alternative models to obtain more than one prediction for the same hour and forecast horizon. To increase forecast accuracy, it is possible to combine the different predictions to obtain a better one or to dynamically select the best one in each time period. Hybrid alternatives based on combining a few selected forecasts can be considered when the number of models is large. One of the most popular ways to combine forecasts is to estimate the coefficients of each prediction model based on its past forecast errors. As an alternative, we propose using multivariate reduction techniques and Markov chain models to combine forecasts. The combination is thus not directly based on the forecast errors. We show that the proposed combination strategies based on dimension reduction techniques provide competitive forecasting results in terms of the Mean Square Error.



Citation: Poncela-Blanco, M.; Poncela, P. Improving Wind Power Forecasts: Combination through Multivariate Dimension Reduction Techniques. *Energies* **2021**, *14*, 1446. <https://doi.org/10.3390/en14051446>

Academic Editor: George Xydis

Received: 20 December 2020

Accepted: 3 March 2021

Published: 6 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: wind power forecasting; combination of forecasts; renewable energy; dimension reduction; principal components; partial least squares; Markov chain model

1. Introduction

Wind power, measured by its installed capacity, has soared in recent years driven by technological development. This trend will continue as many countries have ambitious objectives in wind power. As wind power is intermittent by nature and is a non-dispatchable form of energy, wind power forecasts are needed for large-scale integration of wind turbines into the electrical grid. There have been several attempts to produce prediction tools to forecast wind energy. For recent reviews of the literature about wind power forecasting see, for instance, [1–4]. For a general overview of energy forecasting (load, electricity prices, wind and solar energy), see Hong et al. [5].

As the share of wind energy in the energy mix increases, wind power and wind power forecast errors (that is, the difference between the predicted value and the real production) can adversely impact the operation of the power system [6]. Thus, the operational and economic impact of wind power forecast errors has been analyzed in recent years. For instance, wind power forecast error modeling has a significant impact on wind integration planning studies [7]. Forecast errors of wind and load predictions are also an important factor to consider when determining spinning reserve requirements [8]. There is also abundant research on the impact of wind power forecast in the unit commitment and economic dispatch problem [9].

Even though the benefits of improving wind power forecasts are clear, only a few works deal with forecast combination techniques applied to wind energy. As the relationship between wind and the generated wind power is highly non-linear and time-varying, and given the variety of models available, it is possible to use alternative models based

on different assumptions about the variables involved, or alternatively based on different forecasting techniques, in order to get more than one prediction for the same time and forecast horizon. Then, the combination of forecasts can be used to get the best possible forecast for any particular time and to reduce forecast uncertainty.

The pioneering work about the combination of forecasts is described in Bates and Granger [10]. Since then, many developments have been made. Surveys of this topic are given in [11–14]. In some of these works, the authors have pointed out the lack of optimality of the individual forecasts due to, among other causes, misspecification or misestimation of the models, non-stationarities, breaks, partial information or different loss functions used by forecasters to estimate the model. Thus, independently of the best predictor, and taking into account that all the models are a simplification of a complex reality and that there is no perfect forecast, gains can be always achieved by means of forecast combinations.

There are two main approaches for forecast combination: combining all the predictors contained in the initial set of forecasters or selecting the best one [15]. In the former strategy, the target is to find the best combination of all the initial forecasts. Ideally, the optimal linear combination would outperform the individual forecasts. In reality, it is not clear how close we can get to this ideal situation, and it is then possible that such a combination could be worse than some of the initial forecasts. Additionally, the uncertainty about the performance of the combination increases with non-stationary or small data sets. In the second strategy, the aim is to choose the best predictor in the initial set. This strategy can be interpreted as a dynamic model selection, where the combination tends to put all the weight on the best available predictor. In a non-stationary problem such as wind power forecasting, the best predictor may change over time and the selection of the best predictor for each time step t can lead to instability problems. There is a third hybrid strategy, which is that when the number of predictors is very large, based on regularization, a few selected forecasts are combined [16]. Finally, there is the alternative of imposing the weights without any optimality ambition. The combination giving equal weight to all forecasters is surprisingly very difficult to beat, constituting the so-known forecast combination puzzle as denoted in [17].

Let p_{t+h} and $y_{i,t+h|t}$ denote the output power of a wind farm at time $t+h$ and a wind power forecast for the same time $t+h$ using information up to t , respectively. For electricity market purposes, t is measured in hours and the lead times are $t+h$ with $h = 1, \dots, 36$.

Given a set of forecasters $y_{i,t+h|t}$ with $i = 1, \dots, N$, the goal of the combination is to obtain a single final prediction $\hat{y}_{c,t+h|t}$ which ideally should be at least as precise as the best of the initial forecasters. That is,

$$\hat{y}_{c,t+h|t} = \beta'_t(h)y_{t+h|t} \quad (1)$$

where $y'_{t+h|t} = (y_{1,t+h|t}, \dots, y_{N,t+h|t})$ is the N dimensional vector of the h -period-ahead forecasts at time t and $\beta'_t(h) = (\beta_{1t}(h), \dots, \beta_{Nt}(h))$ is the vector of weighting factors. This vector is calculated for any time step t and forecast horizon h . When necessary, an intercept can be added to Equation (1) to correct a possible bias in the final prediction.

One of the most common ways to estimate the weighting factors is based on the evaluation of past errors of each of the original competing forecasts (see [15,18,19]). Sánchez proposes a two-step adaptive forecast combination procedure to compute alternative combinations, based on a recursive estimation of the mean squared prediction errors [15]. Nielsen et al. use a mean square criterion, and the weights depend on the underlying mean and covariance of the individual forecast errors. The combination can be sub-optimal as mean and covariance have to be estimated. Moreover, this method occasionally produces combined forecasts outside the range of power production that is feasible [18]. Thordarson et al. reformulate the classical regression model for combining forecasts in order to impose restrictions on the combination weights. The weights can be a non-parametric function of particular meteorological variables. The authors maintain that it is not possible to generalize the results because the meteorological variables used may depend on local climatology or on the type of meteorological model used to generate the forecasts [19]. Recently [20],

three different forecast models have been combined applying a constrained least squares regression method (which minimizes the sum of squared errors). The weighting factors are constant (i.e., they are estimated only once), so the combination does not have a good performance at all times t . Forecast combination techniques are also used to solve a different problem, the so-called forecast reconciliation problem (that the sum of the individual wind power forecasts of different wind farms should be equal to the forecast made for the whole set) [21].

At any time t , we can calculate $\hat{e}_{i,t|t-h} = p_t - y_{i,t|t-h}$ with $i = 1, \dots, N$, which are the forecast errors computed as the difference between the wind energy generated at time t and its estimation made at $t - h$ by forecaster i . As h increases, the relation between what happens at t and what we estimated at $t - h$ becomes weaker. Moreover, as h increases, the relation between the prediction of wind energy for $t + h$ and the information we have at t , that is, with the forecast error $\hat{e}_{t|t-h}$ is even weaker. Because of this, we do not compute the weights $\beta_t(h)$ in Equation (1) based on the forecast errors, but rather on capturing the directions that better summarize the information contained in the set of forecasts for each forecast horizon h . The rationale behind multivariate dimension reduction techniques is that it summarizes the information of the initial set of N forecasts to produce a single prediction. There are two groups of dimension-reduction techniques: those that use the information of the variable to forecast when reducing the dimension (supervised learning) and those that do not use it (unsupervised learning). Within the first category, we consider partial least squares (PLS). Another alternative is sliced inverse regression, but it does not perform well for forecast combinations [22]. Within unsupervised learning techniques, we consider principal components (PC) for combining forecasts as in [22]. Later on, PC was applied in the context of interval predictions for electricity price and load [23]. Dynamic factor models decompose the information in a data set (in our case, the different forecasts) into a common and an idiosyncratic part. The common factors would be the part of the forecast shared by all the predictors, while the idiosyncratic part is specific to each predictor and characterizes the disagreement between them. In principle, using the common factors as the linear combinations of the forecasts can give good results as PC for forecast combination [22]. However, PC is preferred as it is a much simpler technique. Other methods, like, for example, Sparse PC, are used only when the number of forecasts is big enough and some of them are discarded, imposing some regularization when reducing the dimension. We do not contemplate the use of independent component analysis (ICA) or non-linear PC as very simple combination schemes (as the simple average of the forecasts) are so far very difficult to beat. This work is based on Poncela et al. [22], considering a different problem and several forecast horizons instead of only one-step-ahead forecasts. Finally, we have also applied another forecast combination strategy assuming a Markov chain to estimate the probabilities of each model forecast to be the best in each time step and forecast horizon [24]. A hybrid approach combining a Bayesian model averaging with ensemble learning was applied recently to the wind-forecasting problem [25], estimating constant weighting factors for each forecast model. In our case, the weighting coefficients of each prediction model are updated at every time step t based on a forgetting factor to highlight the latest real data.

The structure of the paper is as follows: Section 2 introduces the initial set of forecasting models, Section 3 describes the forecast combination procedures we applied in this work; Section 4 describes the data, Section 5 shows the results and, finally, the conclusions are provided in Section 6.

2. Initial Set of Competing Forecasters

The initial set of forecasts is as follows:

NR model: this is a very simple autoregressive model that produces very good results at short forecast horizons, and it is useful in case the input from the meteorological model does not arrive to feed the other models. It is given by

$$p_{t+h|t} = a_h p_t + (1 - a_h) \bar{p}_t \quad (2)$$

where a_h is defined as the correlation coefficient between p_{t+h} and p_t , and \bar{p}_t is the average of the wind energy.

The remaining forecasts come from the time-varying models written in state-space form and optimized in [26]. The state-space model is made up of two equations. The observation equation is defined as:

$$y_t = A_t x_t + e_t \quad (3)$$

where $y_t = p_{t+h}$ is the wind energy to be forecast at time t , $h \in \mathbb{N}^+$ with $h = 1, \dots, 36$ is the forecast horizon; A_t is the observation matrix, $x_t = [x_{0t}, x_{1t}, \dots, x_{mt}]'$ is the state vector; and e_t is the error term with 0 mean and associated variance R .

The state equation is defined as:

$$x_t = \Phi x_{t-1} + \xi_t \quad (4)$$

where Φ is the transition matrix and ξ_t is the error term with 0 mean and associated variance Q . The state vector at time 0 is denoted by $x_0 \sim (x_0, P_0)$, being $b_0 = E(x_0)$ and $P_0 = \text{var}(x)$.

We employ a set of models of increasing complexity. The observation matrix A_t contains the inputs and depends on each specific model. In what follows, we include a brief description of the alternative models:

- M1 model: the observation matrix contains only information of wind power generation in previous periods up to time t :

$$A_t^{(M1)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}] \quad (5)$$

The term p_{t+h-c} attempts to capture the daily cycle. If $h < 24 - k$, then $c = 24$ and p_{t+h-c} is the wind power generated 24 hours before the hour the model is predicting. If $h > 24 - k$, then $c = h + k$. For Sotavento wind farm, $k = 3$.

- M2, M3 and M4 models: these models incorporate the available meteorological wind speed forecast, trying to reproduce the different zones of the wind power curve (relationship between wind and generated energy) as follows:

$$A_t^{(M2)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}] \quad (6)$$

$$A_t^{(M3)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}, \hat{v}_{t+h|t}^2] \quad (7)$$

$$A_t^{(M4)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}, \hat{v}_{t+h|t}^2, \hat{v}_{t+h|t}^3] \quad (8)$$

- M5, M6 and M7 models: these models consider the available meteorological wind speed and wind direction forecast as follows:

$$A_t^{(M5)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}, \sin(\hat{d}_{t+h|t}), \cos(\hat{d}_{t+h|t})] \quad (9)$$

$$A_t^{(M6)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}, \hat{v}_{t+h|t}^2, \sin(\hat{d}_{t+h|t}), \cos(\hat{d}_{t+h|t})] \quad (10)$$

$$A_t^{(M7)} = [1, p_t, p_{t-1}, \dots, p_{t+1-k}, p_{t+h-c}, \hat{v}_{t+h|t}, \hat{v}_{t+h|t}^2, \hat{v}_{t+h|t}^3, \sin(\hat{d}_{t+h|t}), \cos(\hat{d}_{t+h|t})] \quad (11)$$

For each model M1, M2, ..., M7 the size of $A_t^{(Mi)}$ is different, as it is the size of the vectors x_t , and w_t and the matrices Q , R and Φ . R is the variance-covariance matrix of the errors of the observation equation, and Q and Φ characterizes the behavior of the time-varying parameters of the models collected in the state vectors x_t . Diagonal matrices reflect that the parameters do not exhibit any linear relation. A full matrix Φ indicates

that the time-varying parameters depend on their previous states as well as those of the remaining parameters, and a full matrix Q reflects that the parameters are related among them although the correlation enters into the model through the noise covariance matrix. With optimized parameters for each wind farm, we adapt the models to each site. The fitting (learning) of the models is done by estimating the unknown hyperparameters of the model, $\alpha = (x_0, P_0, Q, R, \Phi)$ by maximum likelihood through the Expectation-Maximization algorithm.

Note that we estimate a different model for each forecast horizon h . There are two possibilities to produce h -period-ahead forecasts: direct or iterated forecasts. Direct forecasts are computed using one model for each forecast horizon h , while iterated forecasts are computed from a single model estimated for one-period-ahead forecasts and then iterated forward. There is no consensus about which approach should be preferred, and choosing direct versus iterated approaches is an empirical matter [27]. Theoretically, if we knew the true model, the iterated procedure should produce more efficient forecasts, but direct forecasts are more robust to model misspecification. In the context of forecast combination, we are facing a more realistic situation where the true model is not known and, therefore, there is a set of models to generate forecasts. For wind power, the authors of [26] found that direct forecasts work better for all the wind farms tested.

3. Estimation Methods for Combined Forecasts

3.1. Multivariate Dimension Reduction Techniques

The rationale behind multivariate dimension reduction techniques is to summarize the information of the initial set of N forecasts to produce a single prediction. The final prediction is built on a two-step procedure. In the first step, we build r linear combinations (factors) of the initial set of forecasts as follows:

$$f_{j,t+h|t} = w'_{jt}(h)y_{t+h|t} \text{ with } j = 1, 2, \dots, r \quad (12)$$

where $w_{jt}(h)$ is the vector of weighting factors and $y'_{t+h|t} = (\hat{p}^{nr}_{t+h|t}, \hat{p}^{M1}_{t+h|t}, \dots, \hat{p}^{M7}_{t+h|t})$ is the vector of h -period-ahead forecasts defined in Section 2.

In the second step, we obtain the final prediction as:

$$\hat{p}_{c,t+h|t} = \hat{\beta}_{0t} + \hat{\beta}_{1t}f_{1,t+h|t} + \dots + \hat{\beta}_{rt}f_{r,t+h|t} \quad (13)$$

where $\hat{\beta}_t = (\hat{\beta}_{0t}, \dots, \hat{\beta}_{rt})'$ are the Ordinary Least Squares (OLS) coefficients calculated with the regression

$$p_t = \beta_{0t} + \beta_{1t}f_{1,t|t-h} + \dots + \beta_{rt}f_{r,t|t-h} + u_t \quad (14)$$

where p_t is the wind energy at time step t , $\hat{y}_{c,t+h|t}$ is a real ex-ante forecast as the coefficients $\hat{\beta}_{jt}$ are estimated based on data up to t . Note that we calculate two different weighting vectors: the first one, $w_t(h)$, to compute f_{jt} (linear combinations of the initial set of forecasts), and the second vector, $\beta_t(h)$, to combine the factors and produce the final forecast. As the input set (the competing forecasts) are predictions of the same variable, they could present a high correlation. In this sense, the multivariate reduction dimension techniques extract, in a feasible way, the common information they contain.

We consider two different multivariate dimension reduction techniques to calculate the r factors (f_{1t}, \dots, f_{rt}): partial least squares (PLS) and principal components (PC). The difference between the two techniques is based on the information set used to calculate the factors and the method employed to compute the weighting vector $w_t(h)$. PLS considers the information of the variable to predict p_t , whereas PC does not take into account this information. This means that, in the first case (PLS), $\hat{w}_{jt}(h) = \hat{w}_j(p_h, y_{h|0}, p_{h+1}, y_{h+1|1}, \dots, p_t, y_{t|t-h})$ and in the second case (PC), $\hat{w}_{jt}(h) = \hat{w}_j(y_{h|0}, y_{h+1|1}, \dots, y_{t|t-h})$. The principal components find linear combinations (factors) that explain the variance among forecasts, $var(Y)$,

being Y the matrix that collects all the forecasts while partial least squares look for linear combinations in the direction of $cov(Y, p)$ being p the vector of observations. The PC factors summarize the variability among the initial set of forecasts, while the PLS factors are linear combinations of the initial forecasts that are useful for predicting the wind energy. In what follows, we briefly summarize each technique.

3.1.1. Principal Components

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert an initial set of correlated variables into a new set of uncorrelated variables called principal components. The number of principal components retained should be less than or equal to the number of original variables. The principal components have a decreasing variance value and are orthogonal among them.

Let Y_1, \dots, Y_N be the initial set of variables and Σ_Y its covariance matrix. The ordered principal components are the successive linear combinations $f_i = w_i'Y$, $i = 1, \dots, N$, where w_i is the eigenvector with modulus 1 associated with the i -th eigenvalue of Σ_Y . In our problem, Y contains the alternative forecasts for wind energy, $y_{t+h|t} = (\hat{p}_{t+h|t}^{nr}, \hat{p}_{t+h|t}^{M1}, \dots, \hat{p}_{t+h|t}^{M7})'$.

In the second step of the procedure, we compute the final prediction $\hat{y}_{c,t+h|t}$ with the first r principal components or factors, $f_t = (f_{1,t}, \dots, f_{r,t})'$ and the regression coefficients obtained with Equation (14).

3.1.2. Partial Least Squares

A PLS model tries to find the multidimensional direction in the input variable space (Y) that explains the maximum multidimensional variance direction in the output variable space p [28].

The computation is a sequential process where the correlation of the variable to predict (up to time step t for obtaining true ex-ante predictions) and the initial set of forecasts are used to compute the factors. Let $y_{t|t-h} = (y_{1,t|t-h}, \dots, y_{N,t|t-h})'$ be the vector of h -period-ahead forecasts of the observed variable p_t . The first factor f_{1t} , is obtained projecting the covariance between the variable to predict and its forecasts, in the direction of the forecasts,

$$f_{1t} = \alpha \sum_{i=1}^N cov(y_{i,t|t-h}, p_t) y_{i,t|t-h}, \quad (15)$$

where α is a proportionality constant that depends on the selected normalization. Let $M_{1t}^{p_t}$ be the residuals of the regression of p_t over f_{1t} and $M_{1t}^{y_{i,t|t-h}}$, $i = 1, \dots, N$ the residuals of the N regressions of the variables $y_{i,t|t-h}$ over f_{1t} . To compute the second factor, we use this information, orthogonal to the first factor, and we build the following new linear combination,

$$f_{2t} = \alpha \sum_{i=1}^N cov(M_{1t}^{y_{i,t|t-h}}, M_{1t}^{p_t}) M_{1t}^{y_{i,t|t-h}}, \quad (16)$$

The following component or factor uses the information orthogonal to the previous factors. The procedure is repeated until all factors have been calculated.

3.2. Dynamic Model Averaging

This method was designed to deal with real-time forecasting when there is uncertainty about which will be the best forecast at time t [24]. It combines the initial set of forecasts with a Markov Chain to select the best forecast. The Markov Chain is estimated recursively with a two-step process for each time step t . Let M_1, \dots, M_N be the individual N models, being $L_t = k$ if the real process is driven by model M_k at t . The shift from one prediction model to another is determined by a Markov chain with transition matrix $T(h) = (T_{kl})$ with

$T_{kl}(h) = P[L_{t+h} = l | L_t = k]$, being h the forecast horizon. Letting $\pi_{t|t,l} = P[L_t = l | P_t]$ with $P_t = (p_1, \dots, p_t)$, the prediction equation of the Markov chain probabilities is

$$\pi_{t+h|t,k} = P[L_{t+h} = k | P_t] = \sum_{l=1}^N \pi_{t|t,l} T_{kl} \quad (17)$$

To avoid the need to explicitly estimate the transition matrix T , the proposed algorithm is a recursive procedure where the estimation of the probabilities is based on a forgetting factor. Equation (17) is calculated as

$$\pi_{t+h|t,k} = \frac{\pi_{t|t,k}^\alpha + c}{\sum_{l=1}^N \pi_{t|t,l}^\alpha + c} \quad (18)$$

α is the forgetting factor, whose value is slightly less than one, and c is a small value introduced to avoid numerical instabilities. The updating equation of the Markov chain is:

$$\pi_{t+h|t+h,k} = \frac{w_{t+h,k}}{\sum_{l=1}^N w_{t+h,l}} \quad (19)$$

with

$$w_{t+h,l} = \pi_{t+h|t,l} f_l(y_{t+h} | P_t) \quad (20)$$

being $f_l(p_{t+h} | P_t) \sim N(A_t x_t, R_t)$ the normal distribution density function with information up to time t , obtained from Equations (3) and (4).

The final forecast combination is,

$$\hat{y}_{t+h|t}^{DMA} = \sum_{k=1}^N \pi_{t+h|t,k} y_{k,t+h|t} \quad (21)$$

The process is repeated with each new observation, with $\pi_{0|0,l} = 1/N$ for $l = 1, \dots, N$.

We have applied the same forgetting factor value for all forecast horizons, equal to 0.999.

4. Data

We used data from the Sotavento wind farm to apply the proposed methodology. Sotavento is located in Galicia, Spain [29] and presents a complex orography, which makes wind speed prediction and hence wind power forecast difficult. Sotavento is an experimental wind farm: instead of having identical wind turbines, it consists of 24 wind turbines using 5 different technologies (asynchronous two speeds generator and fixed blade pitch, asynchronous two speeds generator and variable blade pitch, doubly fed induction single-speed generator with variable blade pitch, double fed induction single-speed generator with fixed blade pitch and synchronous generator with variable speed and variable blade pitch), with the objective to be a “showcase” wind farm. Its nominal power is 17.56 MW, and the estimated annual production is 38,500 MWh/year. We have historical measurements of wind power generation and meteorological wind speed and direction forecasts from an numerical weather prediction (NWP) model. The meteorological predictions are daily forecasts of the Advanced Regional Prediction System (ARPS) model at 10 km resolution on its 00Z execution, providing a forecast for the following 72 h. In Sotavento, historical wind power data are obtained every 10 min and then averaged each hour. The data set extends from 17/02/2004 1:00 until 28/04/2005 8:00 h, containing 10471 values. Data are initially divided into two sets. The first year of data, which amount to 8760 data points, were used to estimate and optimize the models. With the remaining data, (1711 points), the first 100 were used for the initial estimation of the covariance matrices of the initial set of forecasters and the remaining 1611 measurements to check the performance of the different forecast combination schemes and to validate the results.

As is usual in the field, power data were divided by the installed capacity and wind speed data were divided by the maximum speed. This produces forecast results that are independent of the wind farm size, and thus we can compare the forecasting performance among different analyses.

5. Results and Discussion

This section shows the results of the combination strategies tested and compares them using several measures of forecast accuracy. As pointed out in [30], the forecasts can be ranked depending on the measure of forecast accuracy chosen. We compare our results according to the following criteria: bias, root mean squared error (RMSE), mean absolute error (MAE) and quantile score (QP) for some quantiles (see [30] for a review of the most popular evaluation metrics for wind power forecasts). To compare the different alternatives for forecast combination, we use the Diebold and Mariano tests for forecast accuracy [31]. In the calculation of the bias, some positive errors are canceled out by negative errors. In terms of the RMSE, all errors are taken into account. Moreover, larger errors have a proportionally greater impact on the RMSE as they are squared. In the electricity system, large wind power forecast errors have a large impact on the system operation and could cause, at certain moments, concerns about the security of supply. As an example, during the cold spell in January 2017, Elia, the transmission system operator in Belgium, had to emit an alert through the European Awareness System and declare that emergency exchanges towards France and the Netherlands were unavailable. This was caused by large errors in their forecast of wind and solar power production (see [32] for more details). Large forecast errors can lead to congestions and inefficient management of power flows. Moreover, forecast errors have to be compensated to keep the system balanced at any time. Typically, imbalances are solved by conventional generators until more storage and demand response will be deployed. It is well known that constant variations of the set point of conventional power plants lower their efficiency, resulting in an increase in the emission factors per MWh and increased maintenance costs due to wear and tear. It is also well known that the maintenance costs of hydro turbines increase when subject to changes in the set points used to balance the electrical grid. In summary, from the network operator perspective and from the environmental point of view, large forecast errors have a larger impact than smaller errors, which can be easily sorted out by netting imbalances. Because of this, we will use the RMSE as the main indicator to compare the results.

For an optimal combination, we do not use an aprioristic number of components or factors. Instead, we carried out the analysis with 1 to 4 components and, a posteriori, we estimated the optimum number in terms of the chosen metric. We kept the number of factors low (from 1 to 4) for several reasons. First, the set of forecasters is not big. Second, each additional factor provides less information. Third, the uncertainty associated with estimating a larger number of components for the weighting vectors can cancel out the positive effect of introducing a new factor.

All numerical results are provided as Supplementary Material in an Excel file.

The set of initial predictors comprises the NR model defined in Equation (2) as well as the multivariate models M1 to M7.

5.1. Benchmark Combinations

Simple benchmark combination techniques were used to compare the results obtained with the proposed schemes. The benchmark models are very simple procedures that, at the same time, present very good practical results. In fact, one of them, the average of forecasts, is hard to beat, a finding that constitutes the “forecast combination puzzle” in the forecasting literature.

The benchmark techniques are: (i) the mean of the initial set of forecasters, (ii) the median and (iii) the ordinary least square (OLS) regression with and without independent term, $y_c(t+h|t) = \alpha_{0t} + X \times \beta_t$ and $y_c(t+h|t) = X \times \beta_t$, respectively.

5.2. Bias

Let $\hat{y}_{t|t-h}^{(k)}$ be the point forecast for wind energy at time t computed at $t - h$ and y_t the observations, with $t = 1, 2, \dots, T$; k denotes that we have different forecasts (the initial set, and the results of the different combination strategies). The bias, defined as

$$\text{Bias}(h) = \frac{1}{T} \sum_{t=1}^T (\hat{y}_{t|t-h}^{(k)} - y_t), \quad (22)$$

measures the ability of a procedure to capture the average level of the target process. The biases for the initial set of forecasts and the combination strategies, are shown in Figure 1. Notice that power data have been divided by the installed capacity of the wind farm, so all values are between 0 and 1. As can be observed in Figure 1, there is no evidence of bias problems in the forecast results. The Supplementary Materials contains the numerical results.

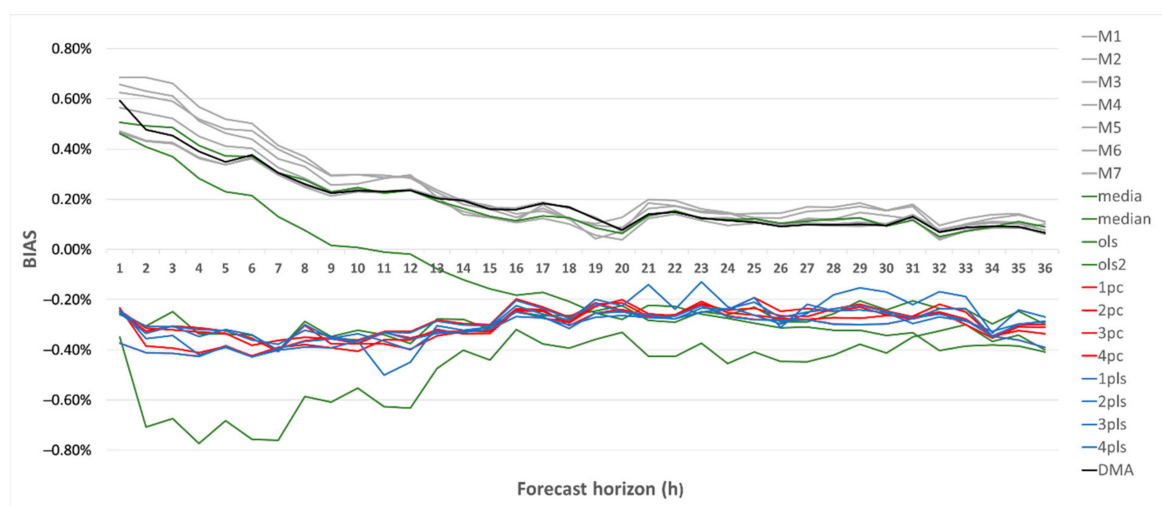


Figure 1. Bias of the different forecasts: (i) initial competing forecasts (M1, M2, M3, M4, M5, M6 and M7 models) represented in grey color, NR has not been included in the figure due to its large bias; (ii) benchmark combination strategies: media, median, ordinary least squares with intercept (denoted as ols) and without intercept (denoted as ols2) represented in green color; (iii) combination based on principal components with one, two, three and four factors, denoted as 1pc, 2pc, 3pc and 4pc, respectively, represented in red color; (iv) combination based on partial least squares with one, two, three and four factors, denoted as 1pls, 2pls, 3pls and 4pls, respectively, represented in blue color; (v) combination based on a Markov Chain denoted as DMA and represented in black color.

5.3. Root Mean Squared Error

The root mean squared error is defined as

$$\text{RMSE}(h) = \frac{1}{N} \sqrt{\sum_{t=1}^N (\hat{y}_{t|t-h}^{(k)} - y_t)^2} \quad (23)$$

Larger errors have a proportionally greater impact on the RMSE as they are squared. This is an important indicator when the main objective is to manage the power grid in real time, i.e., from the network operator perspective.

Figure 2 shows the results of the different combination strategies and the initial set of forecasts in terms of RMSE. It can be seen that the different combination strategies outperform, in terms of RMSE, the results of the initial forecasts. Although the benchmark combinations produce very good results, the combinations based on multivariate reduction techniques (principal components and partial least squares) show a better performance. The Supplementary Material File contains the numerical results.

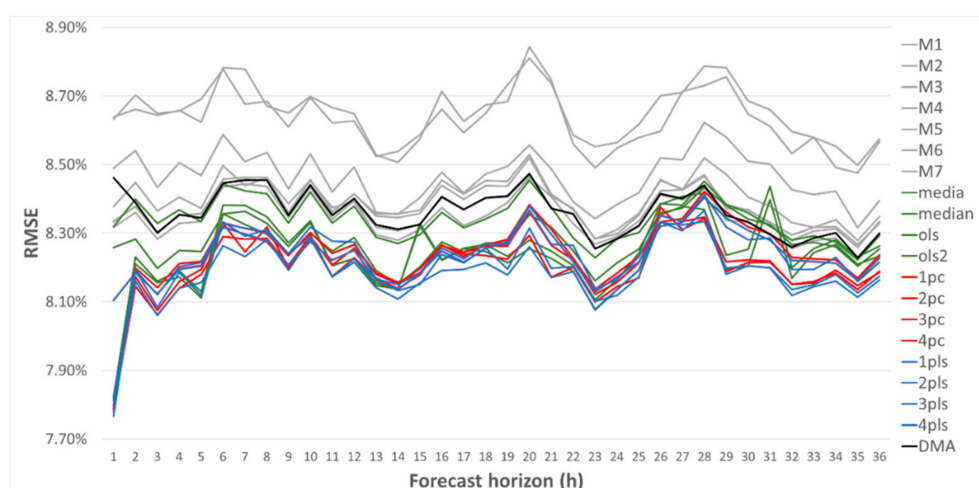


Figure 2. Root Mean Squared Error (RMSE) of the different forecasts: (i) initial competing forecasts (M1, M2, M3, M4, M5, M6 and M7 models) represented in grey color; NR has not been included in the figure due to its large error; (ii) benchmark combination strategies: media, median, ordinary least squares with intercept (denoted as ols) and without intercept (denoted as ols2) represented in green color; (iii) combination based on principal components with one, two, three and four factors, denoted as 1pc, 2pc, 3pc and 4pc, respectively, represented in red color; (iv) combination based on partial least squares with one, two, three and four factors, denoted as 1pls, 2pls, 3pls and 4pls, respectively, represented in blue color; (v) combination based on a Markov Chain denoted as DMA and represented in black color.

5.4. Mean Absolute Error

The mean absolute error (MAE) is computed as

$$MAE(h) = \frac{1}{N} \sum_{t=1}^N |\hat{y}_{t|t-h}^{(k)} - y_t| \quad (24)$$

Figure 3 represents the results in terms of MAE. All the combination strategies present better results than the initial set of forecasts. As the forecast horizon increases, the proposed combination strategies (based on principal components, partial least squares and dynamic moving averaging based on a Markow chain) beat the benchmark strategies.

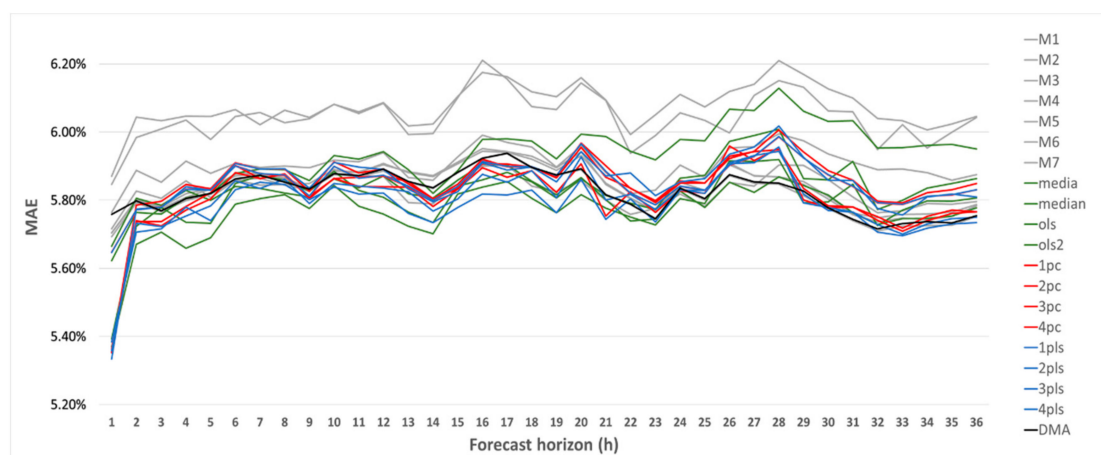


Figure 3. Mean absolute error (MAE) of the different forecasts: (i) initial competing forecasts (M1, M2, M3, M4, M5, M6 and M7 models) represented in grey color; NR has not been included in the figure due to its large errors; (ii) benchmark combination strategies: media, median, ordinary least squares with intercept (denoted as ols) and without intercept (denoted as ols2) represented in green color; (iii) combination based on principal components with one, two, three and four factors, denoted as 1pc, 2pc, 3pc and 4pc, respectively, represented in red color; (iv) combination based on partial least squares with one, two, three and four factors, denoted as 1pls, 2pls, 3pls and 4pls, respectively, represented in blue color; (v) combination based on a Markov Chain denoted as DMA and represented in black color.

5.5. Quantile Scores

Quantile scores are useful when the positive and negative errors have distinct consequences. A different weight is assigned depending on whether the error is positive or negative.

$$QS(p, h) = \frac{1}{N} \sum_{t=1}^N (\hat{y}_{t|t-h}^k - y_t) \times \begin{cases} p & \text{if } (\hat{y}_{t|t-h}^k - y_t) > 0 \\ (1 - p) & \text{if } (\hat{y}_{t|t-h}^k - y_t) < 0 \end{cases} \quad (25)$$

with $0 < p < 1$ being the weighting factor for positive errors. Figures 4 and 5 show the results considering two different p -values, representing two opposite directions of penalising positive and negative errors. Comparing both figures, we can observe how the ranking of the different forecasts changes. With $p = 0.3$, forecast combinations outperform the initial set of forecasts, whereas the opposite is true if $p = 0.7$ (in fact, this can be easily deduced from the fact that $p = 0.3$ implies that $1 - p = 0.7$). Quantile score results for different values of p are provided in the Supplementary Material.

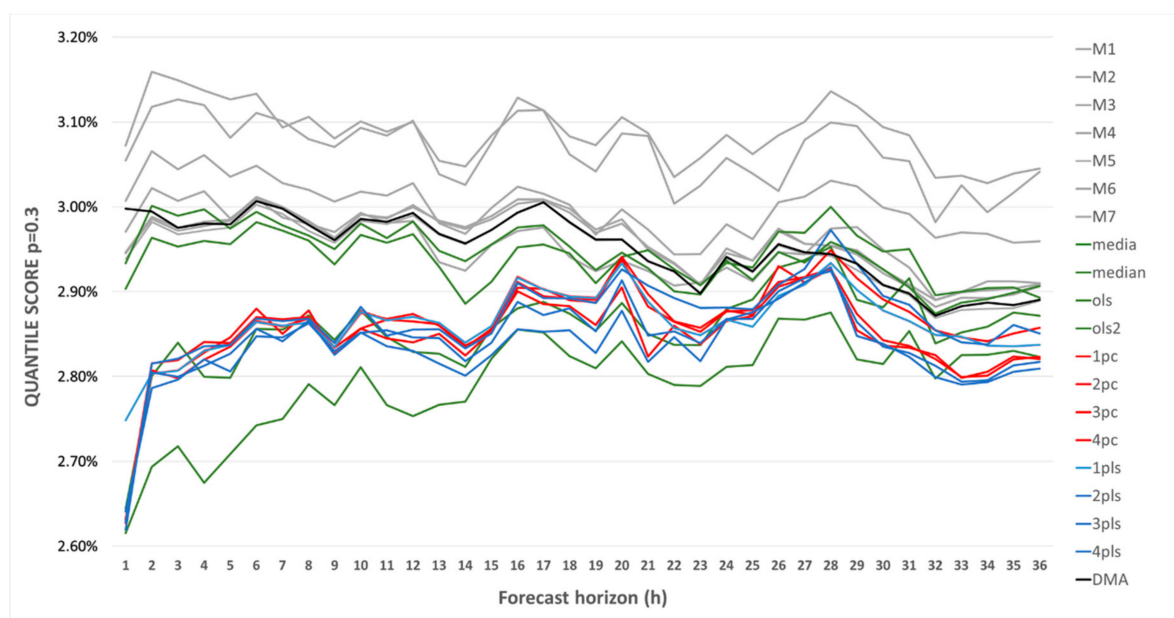


Figure 4. Quantile score of the different forecasts with $p = 0.3$: (i) initial competing forecasts (M1, M2, M3, M4, M5, M6 and M7 models) represented in grey color; NR has not been included in the figure due to its large error; (ii) benchmark combination strategies: media, median, ordinary least squares with intercept (denoted as ols) and without intercept (denoted as ols2) represented in green color; (iii) combination based on principal components with one, two, three and four factors, denoted as 1pc, 2pc, 3pc and 4pc, respectively, represented in red color; (iv) combination based on partial least squares with one, two, three and four factors, denoted as 1pls, 2pls, 3pls and 4pls, respectively, and represented in blue color; (v) combination based on a Markov Chain denoted as DMA and represented in black color.

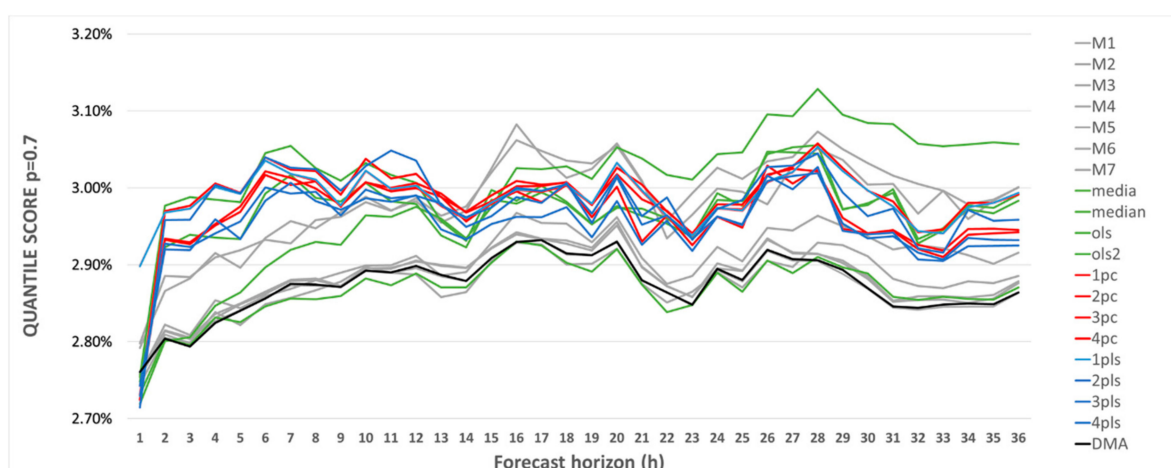


Figure 5. Quantile score of the different forecasts with $p = 0.7$: (i) initial competing forecasts (M1, M2, M3, M4, M5, M6 and M7 models) represented in grey color; NR has not been included in the figure due to its large bias; (ii) benchmark combination strategies: media, median, ordinary least squares with intercept (denoted as ols) and without intercept (denoted as ols2) represented in green color; (iii) combination based on principal components with one, two, three and four factors, denoted as 1pc, 2pc, 3pc and 4pc, respectively, and represented in red color; (iv) combination based on partial least squares with one, two, three and four factors, denoted as 1pls, 2pls, 3pls and 4pls, respectively, and represented in blue color; (v) combination based on a Markov Chain denoted as DMA and represented in black color.

5.6. Tests of Forecast Accuracy

To complete the exposition of results, we have performed the Diebold and Mariano test for forecast accuracy for the mean squared error (MSE). As we have mentioned before, this forecast accuracy measure is of primary interest for network operators.

Let $e_{t+h|t}^{(1)} = p_{t+h} - \hat{y}_{t+h|t}^{(1)}$ and $e_{t+h|t}^{(2)} = p_{t+h} - \hat{y}_{t+h|t}^{(2)}$ be the h -period-ahead forecast errors from two different models and L the loss function such that $L(p_{t+h}, \hat{y}_{t+h|t}^{(i)}) = L(e_{t+h|t}^{(i)})$. The loss differential made at each period t is given by

$$d_t = L(e_{t+h|t}^{(1)}) - L(e_{t+h|t}^{(2)}) \quad (26)$$

The Diebold-Mariano test statistic is computed as

$$S = \frac{\bar{d}}{(\hat{var}(\bar{d}))^{1/2}} = \frac{\bar{d}}{(L\hat{RV}_{\bar{d}}/T)^{1/2}} \quad (27)$$

with $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$ and $L\hat{RV}_{\bar{d}}$ are used to estimate of variance of $\sqrt{T}\bar{d}$ as follows:

$$L\hat{RV}_{\bar{d}} = \gamma_0 + 2 \sum_{j=1}^m \gamma_j, \quad \gamma_j = cov(d_t, d_{t-j}) \quad (28)$$

where the number of lags m used to compute the long-run variance is large enough to take into account the autocorrelation structure of the errors for each forecast horizon.

We summarize the results of the multivariate dimension techniques (PLS and PC) versus OLS with intercept, since this is the most challenging alternative in terms of the MSE according to the results shown in Figure 2. We computed the results for two different sizes of the training sample used to estimate the initial weights of the forecast comparison, $T_0 = 100$ (see Appendix A) and $T_0 = 1000$. In the first case, we leave more data points to increase the significance of the results. For the sake of brevity, we present here a summary of the results for the first case.

Let $e_{t+h|t}^{(1)}$ be h -step-ahead forecast errors from the combination of forecasts performed with OLS and $e_{t+h|t}^{(2)}$ be the h -step-ahead forecast errors from the different forecast combination techniques. Positive values of the test statistic indicate that OLS performs worse than the alternative forecast model. Only for $h = 5$ does OLS yield better results than the best multivariate alternatives. For the remaining forecast horizons, there is always a multivariate dimension technique that outperforms OLS. As our aim is to check if dimension reduction techniques outperform OLS for forecast combination, we compute the p -values for the one side test. Statistical significance in favor of the multivariate techniques is found for medium and longer horizons although, for instance, for $h = 1$, PLS with 2 components yields a p -value of 0.045. Full results are given in Appendix A and in the Supplementary Material File.

6. Conclusions

Wind energy is increasing and gaining importance in electricity generation around the world. Its integration and reliability are essential to ensure the continuity of the supply of electricity. Precise wind energy forecasts are key to efficiently manage the grid. Currently, there are many forecast models available to grid operators or wind farm managers. Due to this plethora of forecast techniques, we have applied different forecast combination strategies based on dimension reduction techniques and Markov chain models, to improve the performance of the final prediction.

One of the most commonly used strategies to combine forecasts is to compute the weighting factors of each individual predictor as a function of its forecast errors: that is, for each forecast horizon h , model i , at time period t , the weighting factor calculated to build the forecast $\hat{y}_{t+h|t}$ is a function of $[e^{(i)}(t|t-h), e^{(i)}(t-1|t-h-1), \dots, e^{(i)}(h|0)]$. As h increases, the relationship between what was estimated at time $t-h$ and what will happen at $t+h$ could be weak, so this strategy may not be adequate for large forecast horizons, since the dynamics of the system may have changed (for example, the system could be in another meteorological regime, and the weighting factor estimated for $t+h$ in t according to the forecast error done at $t-h$ for the forecast at t could be inadequate).

The novelty of this work is that the weights of the initial forecasters are estimated adaptively and are not based on the forecast errors in previous time steps t . Multivariate reduction techniques search for the directions that best summarize the information content of the initial forecasters. In the case of the dynamic model averaging techniques, the combination strategy is based on a Markov chain to estimate the probability of each initial forecast being the best predictor for each time step t and forecast horizon h .

We have analyzed the results of the different combination strategies and compared their performance considering different approaches: bias, RMSE, MAE and quantile scores.

In terms of bias, all forecasts present good results, and it can be considered that all of them are unbiased (bias is less than 0.1% in all cases).

RMSE is important from the network operator perspective. Combination strategies based on multivariate dimension reduction techniques show better results than the initial set of forecasters and the benchmark combination strategy. Partial least squares offers better results than principal components.

To mimic the wind farm owner perspective, we have used MAE and quantile scores, which are among the most common indicators. MAE results show that the proposed combination strategies outperform the initial set of forecasts and most of the benchmark combination strategies in these cases too. Finally, the quantile score results depend on the weight assigned to positive and negative errors.

The combination strategy implemented is a two-step process. First, the input space is reduced, and second, the factors obtained in the first step are combined in a linear regression to obtain the final forecast. In the case of partial least squares, the factors maximize the covariance between the input space data (in our case the set of initial forecasts) and the variable to be predicted, while principal components maximizes the variance of the

initial forecasters without taking into account the variable to be predicted. It is shown that the strategy that employs a more complete set of information presents better results. Concerning the multivariate dimension reduction techniques, we obtain better results increasing the number of factors from one or two factors to three or four factors as the prediction horizon increases, although it should be noted that the difference between three and four factors is minimal. This is because the uncertainty increases with the prediction horizon, and increasing the number of factors adds more information to the prediction albeit at the expense of more estimation uncertainty. When the latter is too large in relation to the new information the additional factor conveys, it does not compensate to add more factors to the combination.

In a nutshell, we have shown the added value of forecast combinations as a simple tool to lower forecast errors. The results obtained using simple combination strategies were promising. We analyzed several combination strategies based on multivariate reduction techniques and Markov chain and have shown that these strategies outperform the initial set of forecasts and the benchmark combination strategies.

Supplementary Materials: Numerical results for the figures and tables are available online at <https://www.mdpi.com/1996-1073/14/5/1446/s1>.

Author Contributions: P.P.: Conceptualization, methodology, formal analysis, validation and writing—review; M.P.-B.: Investigation, data curation, software, resources, validation, writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding: The second author, Pilar Poncela, acknowledges financial support from the Spanish Government, Ministry of Science, contract grant PID2019-108079GB-C22/AEI/10.13039/501100011033.

Conflicts of Interest: The authors declare no conflict of interest. The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Appendix A

Table A1. Diebold-Mariano test statistic of multivariate dimension reduction techniques versus OLS. 1CP means combination with 1 principal component, 2CP means combination considering 2 principal components and so on. Similarly, 1PLS means a combination made with one factor estimated by Partial Least Squares. The forecast horizon is denoted by h .

DM Statistic	1cp	2cp	3cp	4cp	1pls	2pls	3pls	4pls
$h = 1$	−3.7159	1.1840	1.1958	0.8626	−3.7087	1.6902	1.0709	0.7715
$h = 2$	0.2970	0.1378	0.8858	0.5104	0.3000	0.2449	0.7522	0.3931
$h = 3$	0.3396	0.1587	1.0837	0.9195	0.3436	0.3918	1.1222	0.8766
$h = 4$	−0.3878	−0.7882	0.7083	0.2971	−0.3673	−0.6434	0.7749	−0.2492
$h = 5$	−1.7893	−2.1765	−1.5357	−1.9764	−1.7779	−2.1320	−1.1211	−0.2582
$h = 6$	0.3881	0.2705	0.6766	0.2724	0.4066	0.2979	1.0128	0.4478
$h = 7$	0.4495	0.3391	0.5407	0.8249	0.4470	0.3347	0.9226	1.0590
$h = 8$	0.8730	0.4578	0.7455	0.2095	0.8454	0.4099	0.9039	0.6005
$h = 9$	0.9337	0.4085	0.9751	1.0915	0.8890	0.3429	1.4706	0.7488
$h = 10$	0.7722	0.3651	0.6466	0.6005	0.7656	0.1684	0.8080	0.9168
$h = 11$	−0.1682	−0.3580	−0.0238	0.4765	−0.1600	−0.7653	0.4742	0.8379
$h = 12$	−0.5089	−0.9059	−0.7087	−0.0564	−0.5294	−1.1546	0.0410	0.4249

Table A1. Cont.

DM Statistic	1cp	2cp	3cp	4cp	1pls	2pls	3pls	4pls
h = 13	−0.4045	−0.7909	−0.7707	−0.4888	−0.4271	−0.4370	−0.3977	0.4719
h = 14	0.0233	−0.3965	−0.6412	−0.0457	0.0050	−0.3364	0.1733	0.8246
h = 15	1.0000	0.9510	1.0215	1.1444	0.9982	1.0341	1.2327	1.2175
h = 16	−0.6300	−0.9305	−0.8083	−0.7819	−0.6102	−0.7173	−0.3180	1.0665
h = 17	0.5216	0.3052	0.1806	0.3594	0.5295	0.3843	1.4323	1.6687
h = 18	0.2339	0.0146	0.0692	0.4584	0.2363	−0.0295	0.1965	1.0986
h = 19	−0.7129	−0.7234	−0.4579	0.1698	−0.7044	−0.5072	0.8044	1.4090
h = 20	−1.3783	−1.2992	−1.0982	−0.2276	−1.3751	−1.0492	−0.6259	0.4265
h = 21	−0.7241	−0.9480	−0.3064	1.7016	−0.7275	−0.3225	1.1705	1.7556
h = 22	−0.2054	−0.3677	−0.2907	0.1323	−0.1998	−1.0179	0.0802	0.5725
h = 23	−0.4356	−0.5589	−0.3826	0.1958	−0.4223	−0.6843	0.1268	0.9300
h = 24	−0.1239	−0.3138	0.1424	0.3630	−0.1151	0.0013	1.1265	0.8568
h = 25	0.3267	0.1447	0.4578	1.1951	0.3326	0.1770	1.1024	1.4629
h = 26	0.8601	0.7073	0.5961	0.1868	0.8623	0.5163	1.3221	1.4942
h = 27	1.0330	0.9021	1.3457	1.4451	1.0337	0.9988	1.5777	1.2072
h = 28	0.1771	−0.0562	1.1261	1.8376	0.1719	0.0884	1.5861	2.1658
h = 29	−1.3020	−1.5227	0.3642	1.0578	−1.3069	−1.0950	0.7993	1.3808
h = 30	−0.7225	−0.9236	0.5173	0.6388	−0.7277	−0.4551	0.7921	0.8180
h = 31	0.6340	0.5726	0.9200	0.8967	0.6331	0.6548	0.9798	0.9203
h = 32	−0.3126	−0.4661	0.8152	0.7581	−0.3192	0.0897	1.0469	1.3640
h = 33	0.3888	0.3250	1.0817	1.1045	0.3854	0.6721	1.1932	1.3063
h = 34	0.8248	0.7178	1.1301	1.0400	0.8211	0.6441	1.1786	1.4810
h = 35	0.5717	0.5235	0.9026	0.8285	0.5697	0.6596	1.0351	1.2603
h = 36	0.4370	0.3216	0.9963	1.0584	0.4320	0.5470	1.2203	1.3503

References

1. Foley, A.M.; Leahy, P.G.; Marvuglia, A.; McKeogh, E.J. Current Methods and Advances in Forecasting of Wind Power Generation. *Renew. Energy* **2012**, *37*, 1–8. [\[CrossRef\]](#)
2. Giebel, G.; Kariniotakis, G. Wind power forecasting—a review of the state of the art. In *Renewable Energy Forecasting*; Woodhead Publishing: Cambridge, UK, 2017; pp. 59–109. ISBN 978-0-08-100504-0.
3. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies* **2020**, *13*, 3764. [\[CrossRef\]](#)
4. Tian, Z. A State-of-the-Art Review on Wind Power Deterministic Prediction. *Wind Eng.* **2020**, 0309524X20941203. [\[CrossRef\]](#)
5. Hong, T.; Pinson, P.; Wang, Y.; Weron, R.; Yang, D.; Zareipour, H. Energy Forecasting: A Review and Outlook. *IEEE Open Access J. Power Energy* **2020**, *7*, 376–388. [\[CrossRef\]](#)
6. Ahmed, A.; Khalid, M. A Review on the Selected Applications of Forecasting Models in Renewable Power Systems. *Renew. Sustain. Energy Rev.* **2019**, *100*, 9–21. [\[CrossRef\]](#)
7. Hodge, B.-M.S.; Ela, E.G.; Milligan, M. Characterizing and Modeling Wind Power Forecast Errors from Operational Systems for Use in Wind Integration Planning Studies. *Wind Eng.* **2012**, *36*, 509–524. [\[CrossRef\]](#)
8. Ortega-Vazquez, M.A.; Kirschen, D.S. Estimating the Spinning Reserve Requirements in Systems with Significant Wind Power Generation Penetration. *IEEE Trans. Power Syst.* **2009**, *24*, 114–124. [\[CrossRef\]](#)
9. Ellahi, M.; Abbas, G.; Khan, I.; Koola, P.M.; Nasir, M.; Raza, A.; Farooq, U. Recent Approaches of Forecasting and Optimal Economic Dispatch to Overcome Intermittency of Wind and Photovoltaic (PV) Systems: A Review. *Energies* **2019**, *12*, 4392. [\[CrossRef\]](#)
10. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *J. Oper. Res. Soc.* **1969**, *20*, 451–468. [\[CrossRef\]](#)
11. Clemen, R.T. Combining Forecasts: A Review and Annotated Bibliography. *Int. J. Forecast.* **1989**, *5*, 559–583. [\[CrossRef\]](#)

12. Diebold, F.; Lopez, J. *Forecast Evaluation and Combination*; National Bureau of Economic Research, Inc.: Cambridge, MA, USA, 1996.
13. de Menezes, L.M.; Bunn, D.W.; Taylor, J.W. Review of Guidelines for the Use of Combined Forecasts. *Eur. J. Oper. Res.* **2000**, *120*, 190–204. [[CrossRef](#)]
14. Newbold, P.; Harvey, D.I. Forecast Combination and Encompassing. In *A Companion to Economic Forecasting*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2004; pp. 268–283. ISBN 978-0-470-99643-0.
15. Sánchez, I. Adaptive Combination of Forecasts with Application to Wind Energy. *Int. J. Forecast.* **2008**, *24*, 679–693. [[CrossRef](#)]
16. Diebold, F.X.; Shin, M. Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian LASSO and Its Derivatives. *Int. J. Forecast.* **2019**, *35*, 1679–1691. [[CrossRef](#)]
17. Stock, J.H.; Watson, M.W. Combination Forecasts of Output Growth in a Seven-Country Data Set. *J. Forecast.* **2004**, *23*, 405–430. [[CrossRef](#)]
18. Nielsen, H.A.; Nielsen, T.S.; Madsen, H.; Pindado, M.J.S.I.; Marti, I. Optimal Combination of Wind Power Forecasts. *Wind Energy* **2007**, *10*, 471–482. [[CrossRef](#)]
19. Thordarson, F.Ö.; Madsen, H.; Nielsen, H.A.; Pinson, P. Conditional Weighted Combination of Wind Power Forecasts. *Wind Energy* **2010**, *13*, 751–763. [[CrossRef](#)]
20. Kim, Y.; Hur, J. An Ensemble Forecasting Model of Wind Power Outputs Based on Improved Statistical Approaches. *Energies* **2020**, *13*, 1071. [[CrossRef](#)]
21. Bai, L.; Pinson, P. Distributed Reconciliation in Day-Ahead Wind Power Forecasting. *Energies* **2019**, *12*, 1112. [[CrossRef](#)]
22. Poncela, P.; Rodríguez, J.; Sánchez-Mangas, R.; Senra, E. Forecast Combination through Dimension Reduction Techniques. *Int. J. Forecast.* **2011**, *27*, 224–237. [[CrossRef](#)]
23. Maciejowska, K.; Nowotarski, J.; Weron, R. Probabilistic Forecasting of Electricity Spot Prices Using Factor Quantile Regression Averaging. *Int. J. Forecast.* **2016**, *32*, 957–965. [[CrossRef](#)]
24. Raftery, A.E.; Kárný, M.; Ettler, P. Online Prediction under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics* **2010**, *52*, 52–66. [[CrossRef](#)] [[PubMed](#)]
25. Wang, G.; Jia, R.; Liu, J.; Zhang, H. A Hybrid Wind Power Forecasting Approach Based on Bayesian Model Averaging and Ensemble Learning. *Renew. Energy* **2020**, *145*, 2426–2434. [[CrossRef](#)]
26. Poncela, M.; Poncela, P.; Perán, J.R. Automatic Tuning of Kalman Filters by Maximum Likelihood Methods for Wind Energy Forecasting. *Appl. Energy* **2013**, *108*, 349–362. [[CrossRef](#)]
27. Marcellino, M.; Stock, J.H.; Watson, M.W. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *J. Econom.* **2006**, *135*, 499–526. [[CrossRef](#)]
28. Wold, H. *Estimation of Principal Components and Related Models by Iterative Least Squares*; Krishnaiah, P.R., Ed.; Multivariate Analysis; Academic Press: New York, NY, USA, 1966; pp. 391–420.
29. Sotavento Sotavento Wind Farm. Available online: <https://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/> (accessed on 12 December 2020).
30. Messner, J.W.; Pinson, P.; Browell, J.; Bjerregård, M.B.; Schicker, I. Evaluation of Wind Power Forecasts—An Up-to-date View. *Wind Energy* **2020**, *23*, 1461–1481. [[CrossRef](#)]
31. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
32. ENTSO-E. *Managing Critical Grid Situations—Success and Challenges*. ENTSO-E Report of the January 2017 Cold Spell; ENTSO-E AISBL: Brussels, Belgium, 2017.